Review of egusphere-2025-1772

General comments

The study by Li et al. aims at presenting a method that allows the automated mapping of glacier extents in a challenging region using Google Earth Engine at high temporal resolution, up to annually from Sentinel-2. They used the results to consider area changes when calculating glacier mass balance. If the mapping had worked, this would have been an important study to improve related results also for other regions in the world. Unfortunately, the outcome of the mapping is not useful for any assessment. In this regard I want to acknowledge that the authors have provided the results of their mapping effort in the supplemental material. Without this, my evaluation would have been different as the paper is otherwise well written and the idea to just use all data available and combine it for the best possible result is fine. However, the largely arbitrary area changes from dataset to dataset (for individual glaciers partly larger than 50% from year to year) are obvious and glaciologically impossible. The authors mention that there are the usual problems with debris cover, clouds and shadow, but they have seemingly not recognized how large and arbitrary the variability is and that their method does not produce meaningful results.

Neither the unrealistic area increase by about 500 km^2 (estimated from Fig. 8, numbers for individual years are not provided) from 2019 to 2022 (naming it as a 'consistent decline' in L387), nor the sudden strong increase from 2016 to 2017 is discussed or considered as unrealistic. Instead, the authors correlate glacier elevation changes (wrongly labelled as 'Glacier Thickness' in Fig. 8) with glacier area changes as they assume there is a correlation (L407) and think that the correlation can be used as a validation (L413) of their (wrong) glacier areas. In fact, area changes are mostly driven by the ice thickness distribution along the glacier perimeter (thus depending on the shape of the glacier cross-profile) and are a longer-term response to changes in flow dynamics (glaciers have a response time). Hence, also the follow-on analysis is a bit strange. In this regard, it is also unclear to me why the authors rely on results from Cryosat and ICESat (with their diverse range of issues) for such small glaciers instead of the Hugonnet et al. (2021) dataset that is widely used? As this dataset is not even mentioned in the comparison Table 1, I wonder why. Is the dataset too bad in quality?

The very short results section (it has just 14 lines) mentioned the Kappa coefficient and overall accuracy along with three images showing outline overlays. I am aware that these statistical accuracy measures are frequently used in remote sensing studies to present the accuracy, but in my view they can be the result of anything and do not allow to obtain meaningful conclusions about the 'robustness and reliability of the classification approach'. At least for glaciers they do not work, as nicely confirmed here by the largely arbitrary results of the glacier mapping. The quality of mapped glacier extents can be shown by a) outline overlays and b) the sum of commission and omission errors (false positives and false negatives) divided by the common area. But as the former have been removed by the masking with RGI 7.0, I am unclear if the measures can be used here at all?

I see missing debris-covered glacier parts and that large regions in shadow are sometimes missing. Hence, intensive manual editing would be required before resulting outlines could be used for change assessment. The statement that glaciers 'are well identified' (L377) seems misleading in this regard. Although the year 2000 dataset is likely the most complete regarding shadow and debris mapping, showing a region where the method does not work and discussing it would have been more helpful. One can see the problems of the classification a little bit for the 2022 outlines of the right glacier in the upper left panel [please name them

properly a), b) and c)] of Fig. 8, but the image is very dark (what about some contrast stretching?), the lines are hard to see (also the red outlines on a reddish background in the insets of Fig. 7 are barely visible) and the wrong mapping results are not really discussed.

A final major point of concern is the general set-up of the study. First, the elevation change datasets are introduced in the Discussion Section 5.2 rather than in Sections 2.2 and 3. Their description is thus very short and the processing method unclear (e.g. how has the radar penetration into snow been corrected?). This can likely easily be adjusted. The motivation to determine annual area changes is mentioned, but not critically discussed. Even when the resulting glacier outlines would have been correct, a one-pixel (two pixels for debris-covered regions) uncertainty at 30 m resolution relates to a 30 and 60 m location uncertainty of the outline. With an assumed annual terminus retreat of 5 to 10 m / year (much less around the perimeter), one has to wait several years before new outlines make sense compared to uncertainties. But here the mapped termini could be wrong by several km, so change assessment is not an option. To make my major objection of the high variability in the mapping results clearer, I have added all glacier maps (setting the no data value to 0 before) and received a very colourful picture. On the last two pages of this review, I show a few examples for illustration. If the mapping had been correct, colours should only appear near the terminus and around the perimeter. As a note, this is just the result of a simple adding without a timeline, not revealing the partly strong year-to-year jumps in mapped glacier areas.

In conclusion, this brute-force mapping using sophisticated image processing without a sufficient understanding of the mapped subject (glaciers) and how it should change over time is not recommended. When being harsh, I would ask the authors to please first learn the basics about glaciers and how they work, then proceed with the user needs (are annual updates really required?) and then do the mapping. A bit less harsh I would ask the authors to first get the mapping right for one year before applying it to several years. When glacier area changes are mostly due to changes in the mapping results rather than real changes, there is no need to perform change assessment. In my view, it is possible to publish a study revealing that a method has not worked. However, in this case an honest discussion and illustration of the problems is required to be helpful for future studies. Concluding that this study provides ‘effective support for future glacier inventories’ (L482) is in my view highly misleading.

Specific comments

I do not comment here on all details of the study, but include some general remarks:
Providing area changes in km^2 and mass changes in Gt (sure to use 900 rather than 850 kg/m^3 in this region?) are not useful as they are incomparable across regions. In future studies, please give relative area changes in % and the related change rates per year for area and specific mass changes per unit area and year for mass balance (and please do not use the latter to ‘validate’ the former, this makes glaciologically little sense).

Please carefully check text errors. Often spaces are missing or units are wrong (area in km instead of km^2, volume in m instead of m^3). Also the citation style is strange. For example, in L398 it is written ‘Ye et al. (Quinghua, 2019 ...)reported ...’ So is it now Ye or Quinghua and why is it first et al. and then without et al.? In the reference section it is actually Quinhua, Y.E., again different. Correct would have been to write: ‘The datasets by Qunighua (2019 and 2020) reported ...’ or in L409: ‘The results of Jakob et al. (2021)’ … . As a small note, the References Section becomes more readable when indenting the text from the second line a bit, making it ‘hanging’.

Figure captions: I suggest inserting a . or : after the figure number, e.g. ‘Figure 1: Study area’
Table 1: I think the brackets around the authors of the cited studies are not required.

L261: Figure 4: The blue and red lines and squares are difficult to see against the dark background. Also the annotations and legends of the insets are partly hard to see. It needs also to be explained what is what. Just writing in the text that types can be clearly distinguished is a bit thin. The same applies to all panels in Fig. 5. The panels are too small, the legends are unreadable and it is unclear what is what.

L306: I am also a bit unclear what Fig. 5 should tell me? That many datasets have been used and none of them shows glaciers clearly?

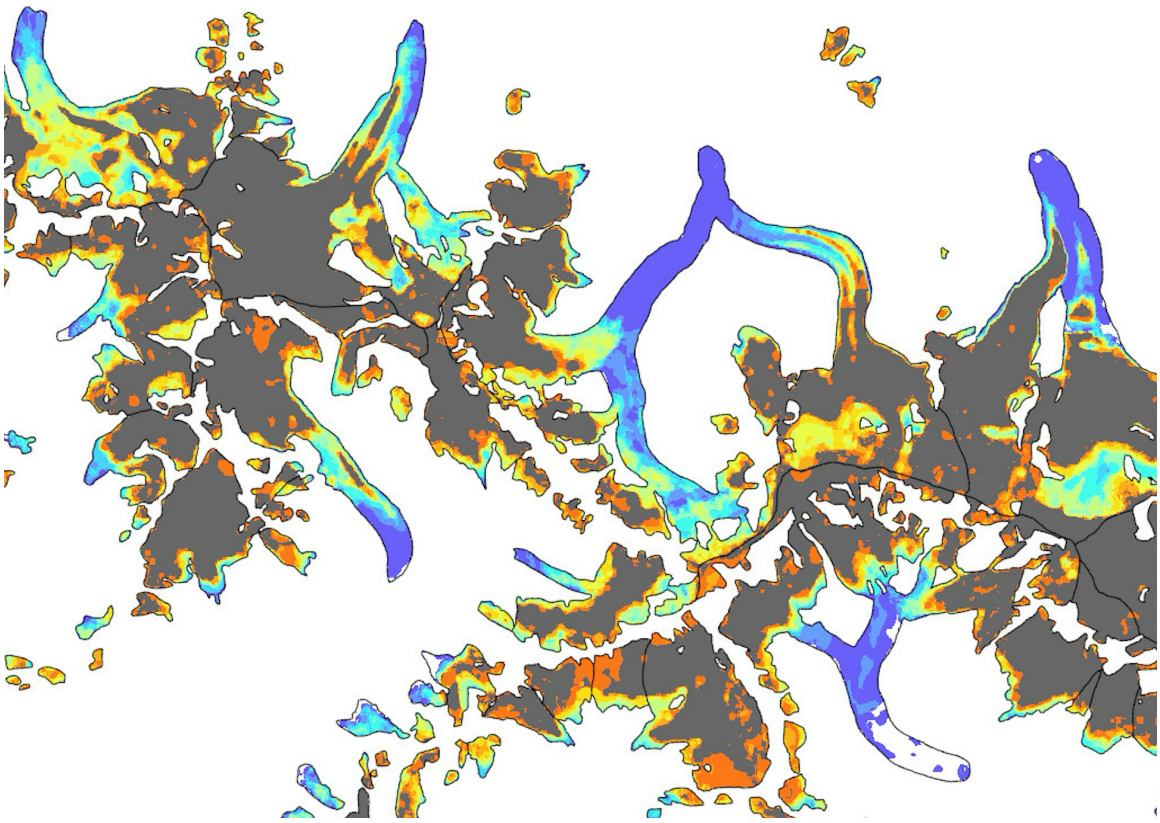
L390: Incomplete caption. Please note that sudden area gains as shown in Figure 5 (and 6) are glaciologically not possible. This is not how glaciers work.

L414: Figure 9b: This comparison makes glaciologically no sense.

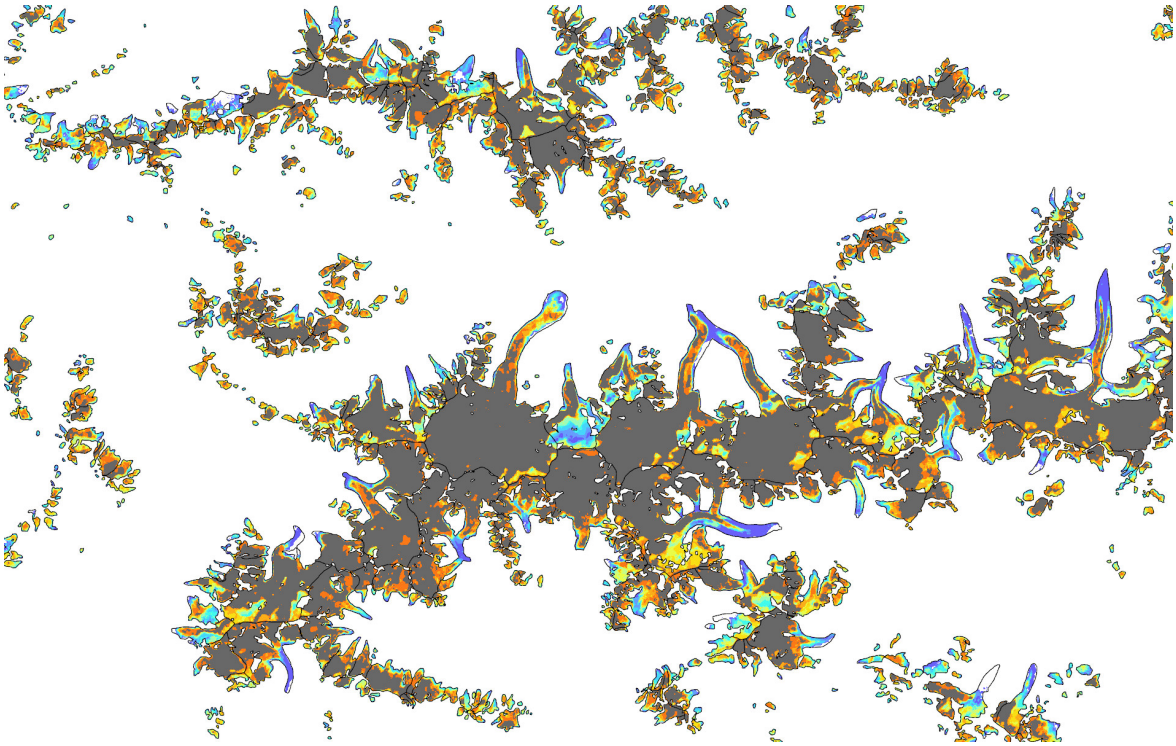
L470: This is correct, but the NDSI has been shown to be very sensitive to path radiance in the green band, creating problems with ice in shadow. Additional classification problems are introduced when using the analysis ready reflectance datasets instead of the raw data, which allow for a better separation of details when the SNR is low.

L471: As far as I can see it, most gaps are due to not mapping debris-covered glacier parts.

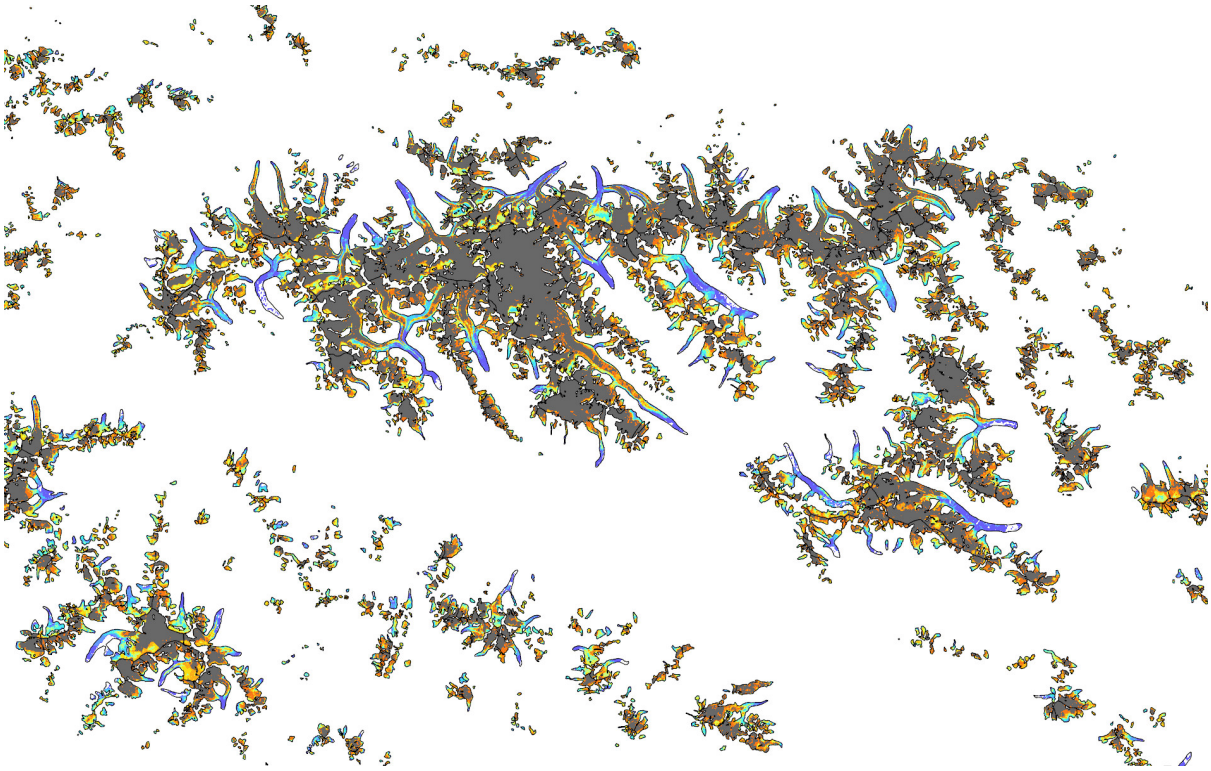
As mentioned above, the illustrations below of the highly variable mapping results presented in this study. Colours should only appear near the glacier termini and (to a lesser degree) along the perimeter. Colours denote regions that have been mapped in all years (dark grey) and by 10 (red) to one (blue) scene. As most debris covered tongues are bluish, they have only been mapped by a few scenes. Black outlines show glacier extents from RGI 7.0.



Example 1



Example 2



Example 3: No colour should be visible at this scale.